

YM500v2: a small RNA sequencing (smRNA-seq) database for human cancer miRNome research

Wei-Chung Cheng^{1,†}, I-Fang Chung^{2,†}, Cheng-Fong Tsai^{2,3}, Tse-Shun Huang⁴,
Chen-Yang Chen^{2,3}, Shao-Chuan Wang², Ting-Yu Chang³, Hsing-Jen Sun², Jeffrey
Yung-Chuan Chao^{5,6}, Cheng-Chung Cheng⁷, Cheng-Wen Wu^{5,8,9} and
Hsei-Wei Wang^{2,3,5,10,11,*}

¹Research Center for Tumor Medical Science, China Medical University, Taichung 40402, Taiwan, ²Institute of Biomedical Informatics, National Yang-Ming University, Taipei 11221, Taiwan, ³VGH-YM Genomic Research Center, National Yang-Ming University, Taipei 11221, Taiwan, ⁴Institute of Engineering in Medicine, University of California, San Diego, La Jolla, CA 92093, USA, ⁵Institute of Clinical Medicine, Medical College, National Yang-Ming University, Taipei 11221, Taiwan, ⁶Department of Radiation Oncology, Taichung Veterans' General Hospital, Taichung 40705, Taiwan, ⁷Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defence Medical Center, Taipei 11490, Taiwan, ⁸Institute of Biomedical Science, Academia Sinica, Taipei 11529, Taiwan, ⁹Institute of Biochemistry and Molecular Biology, National Yang Ming University, Taipei 11221, Taiwan, ¹⁰Institute of Microbiology and Immunology, National Yang-Ming University, Taipei 11221, Taiwan and ¹¹Department of Education and Research, Taipei City Hospital, Taipei 10341, Taiwan

Received September 15, 2014; Accepted October 29, 2014

ABSTRACT

We previously presented YM500, which is an integrated database for miRNA quantification, isomiR identification, arm switching discovery and novel miRNA prediction from 468 human smRNA-seq datasets. Here in this updated YM500v2 database (<http://ngs.ym.edu.tw/ym500/>), we focus on the cancer miRNome to make the database more disease-orientated. New miRNA-related algorithms developed after YM500 were included in YM500v2, and, more significantly, more than 8000 cancer-related smRNA-seq datasets (including those of primary tumors, paired normal tissues, PBMC, recurrent tumors, and metastatic tumors) were incorporated into YM500v2. Novel miRNAs (miRNAs not included in the miRBase R21) were not only predicted by three independent algorithms but also cleaned by a new *in silico* filtration strategy and validated by wetlab data such as Cross-Linked ImmunoPrecipitation sequencing (CLIP-seq) to reduce the false-positive rate. A new function 'Meta-analysis' is additionally provided for allowing users to identify real-time differentially expressed miRNAs and arm-switching events according to customer-defined sample groups and dozens of clinical criteria tidying up by proficient clinicians.

Cancer miRNAs identified hold the potential for both basic research and biotech applications.

INTRODUCTION

MicroRNAs (miRNAs), a small RNA species of ~22 nt in length, control gene expression and are involved in the regulation of a variety of biological processes including development, differentiation, cell proliferation, metabolism and inflammation, as well as in human diseases (1). It is estimated that more than 60% of human protein-coding genes contain miRNA target sites with their 3' UTR regions (2). miRNAs can function as oncogenes or tumor suppressors, as do miR-21 and let-7, respectively, and improved knowledge of them has been one of the defining developments in cancer research over the past decade. The abnormal expression of miRNAs and dysregulation of factors that regulate miRNAs contribute to the progression of tumor (3). Owing to the stability of miRNAs within clinical samples, miRNAs have been regarded as potential prognostic indicators in cancers (4) and as biomarkers for cancer classification (5–8). Many miRNA signatures have been proposed for patient prognosis and clinical response and are being investigated in clinical trials (9). In addition, miRNAs have emerged as therapeutic targets for cancer treatments. The function of miRNAs can be efficiently and specifically inhibited by artificial miRNA mimics or antagomirs, supporting their potentials as novel therapy tactics for diseases (9–

*To whom correspondence should be addressed. Tel: +886 2 28267109; Fax: +886 2 28122880; Email: hwwang@ym.edu.tw

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

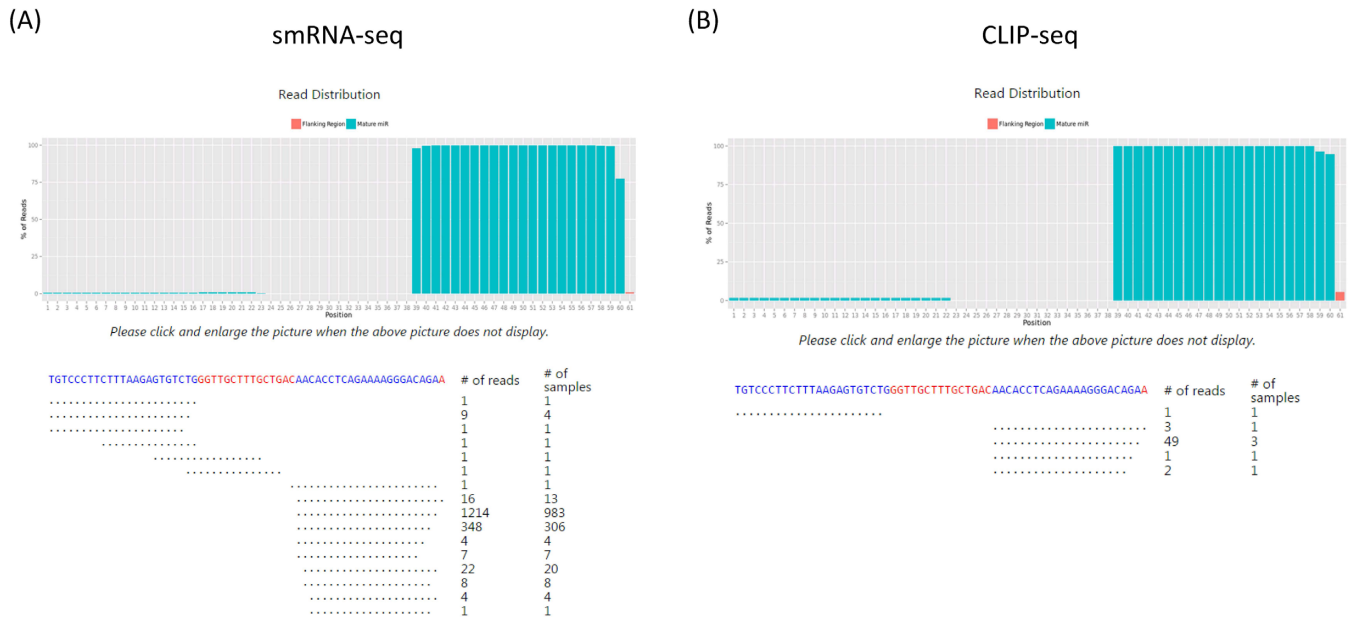


Figure 1. A representation of example of ‘novel miRNA’, indicating the alignment results of the reads from smRNA-seq (A) and CLIP-seq (B).

13). MRX34, a synthetic miR-34a mimic loaded in liposome, is the first miRNA-based therapy and has entered the clinics for cancer therapy in 2013 (14).

In the past few years, because of the increasing use of next generation sequencing (NGS), enormous amounts of small RNA sequencing data have been generated from large-scale cancer projects such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). These data help researchers to investigate the roles of miRNAs in cancer progression. However, translating such massive amounts of data into information that can be easily interpreted and accessed remains a challenge.

Previously, we developed YM500 (15), a database that includes integrated pipelines for miRNA quantification, isomiR identification, arm switching discovery and novel miRNA prediction from smRNA-seq. YM500 provides researchers with integrated miRNA-related information with various graphical visualization pages from 468 human and 141 mouse smRNA-seq datasets *via* a user-friendly web interface. No particular biological question or disease was emphasized in YM500. Here we present YM500v2, an updated version of the database. In addition to including more smRNA-seq results (>8000 TCGA cancer-related datasets), we focus on human cancer research in this version. Other new features also include novel miRNA identification and validation: novel miRNAs, i.e. those not included in miRBase R21, were identified from thousands of samples, and a new strategy was developed for filtering out false-positive discoveries. Especially, dozens of CLIP-seq datasets are provided as experimental evidence for *in silico* predicted novel miRNAs. Regarding isomiRs, two new statistical charts are provided for a specific isomiR. YM500v2 also contains a new function ‘Meta-analysis’, which allows researchers to identify differentially expressed miRNAs and arm-switching events in two customer-defined groups of samples.

DATA COLLECTION AND PRE-PROCESSING

There are 8105 smRNA-seq datasets from TCGA, including those of primary tumors, paired normal tissues, peripheral blood mononuclear cell (PBMC), recurrent tumors and metastatic tumors, incorporated into YM500v2 (Supplemental Table S1). Raw data were downloaded from CGHub (<https://cghub.ucsc.edu/>) and were pre-processed by in-house scripts. To avoid the disturbance of reads with poor quality, we filtered out reads that do not meet the following criterion: Phred Score >30 in >90% sequence. About 30–55% of reads were excluded in this step, but millions of reads still remained for each dataset. As shown in Supplemental Figure S1, this filtration step dramatically improves data quality, which is important for novel miRNA and isomiR identification. The other pre-processing steps were the same as those followed in our previous study (15). Clinical information of each sample was manually curated by proficient clinicians based on clinical data obtained in TCGA. Each sample was annotated with 35 clinical characteristics.

NOVEL miRNA PREDICTION

In this version, 22 CLIP-seq datasets are incorporated as experimental evidence for the putative novel miRNAs, and a new strategy was developed to reduce the false-positive rate. It has been noted that there are numerous advantages in conducting novel miRNA prediction by performing a single analysis of the pooled data (16–18). After pre-processing, we pooled 8105 and 468 smRNA-seq datasets from TCGA and the Gene Expression Omnibus (human datasets in YM500), respectively, comprising approximately 11.1 billion reads from 193 million unique sequences. When analyzing the sequences, we found that there are 76.7% of sequences with only one read in all the samples and 0.41% of sequences with more than 100 reads in ~8600 samples

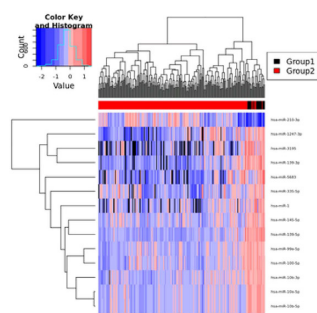
Meta-Analysis Results (ID: 257639549)

Sample Information

	Group One	Group Two
Dataset	Breast_invasive_carcinoma;	Breast_invasive_carcinoma;
Count	19	211
Sample Type	Solid Tissue Normal;	Primary Solid Tumor;
Clinical Condition	breast_carcinoma_estrogen_receptor_status:Negative;	breast_carcinoma_estrogen_receptor_status:Negative;

Differentially Expressed miRNAs

Heatmap of differentially expressed miRNAs



Heatmap of differentially expressed miRNAs: [Download](#) (pdf)

Expression matrix of significant miRNAs: [Download](#) (zip)

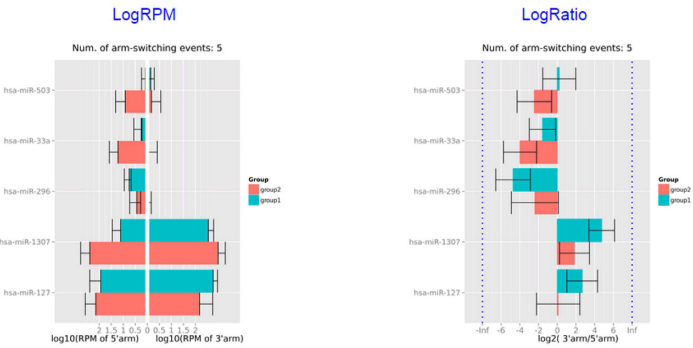
Statistical results for all miRNAs: [Download](#) (csv)

Statistical results for significant miRNAs: [Download](#) (csv)

Statistical results for significant miRNAs (at most top 20 as shown below)

miR Name	Mean expression	Group 1 Mean expression	Group 2 Mean expression	foldChange	log2FoldChange	pval	padj
hsa-miR-1	6.58848	31.46977	4.34799	0.13816	-2.85555	1.0E-5	0.00384
hsa-miR-100-5p	7306.70768	29363.16313	5320.58136	0.1812	-2.46435	0.00011	0.02395
hsa-miR-10a-5p	97715.05446	412970.24009	69327.14675	0.16787	-2.57455	2.0E-5	0.0071
hsa-miR-10b-3p	18.48414	65.6664	14.2355	0.21679	-2.20566	9.0E-5	0.02172
hsa-miR-10b-5p	102673.28055	444985.03063	71848.99974	0.16146	-2.63072	1.0E-5	0.00506
hsa-miR-1247-3p	86.88062	323.15583	65.60465	0.20301	-2.30036	0.00017	0.0339
hsa-miR-139-3p	32.62337	219.47765	15.79763	0.07198	-3.79629	0	0

Arm-switching miRNAs



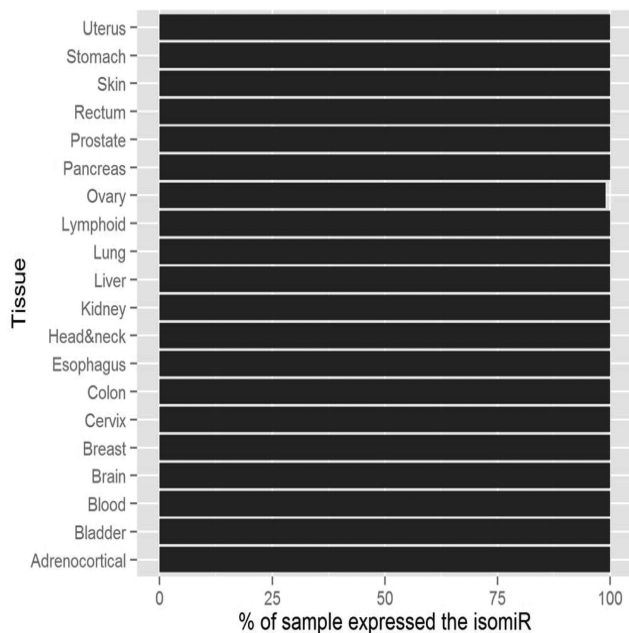
Statistical results for arm-switching miRNAs: [Download](#) (csv)

Statistical results for arm-switching miRNAs (at most top 20 as shown below)

miR_species	RPM_3p_group1	RPM_3p_group2	RPM_5p_group1	RPM_5p_group2	logratio_group1	logratio_group2	Pvalue
hsa-miR-1307	1114.15314	392.91563	362.22608	17.71415	1.84438	4.75528	0
hsa-miR-33a	1.58873	0.65912	23.29738	2.25226	-3.98306	-1.58077	0
hsa-miR-127	327.02674	596.16095	230.21343	150.42674	0.10003	2.66754	0
hsa-miR-503	2.2275	1.18854	12.56933	1.06093	-2.45333	0.214	0
hsa-miR-296	0.95924	0.32459	3.27657	5.80752	-2.38993	-4.72393	2.0E-5

Figure 2. The screen shot of the results for ‘Meta-Analysis’.

(A)



(B)

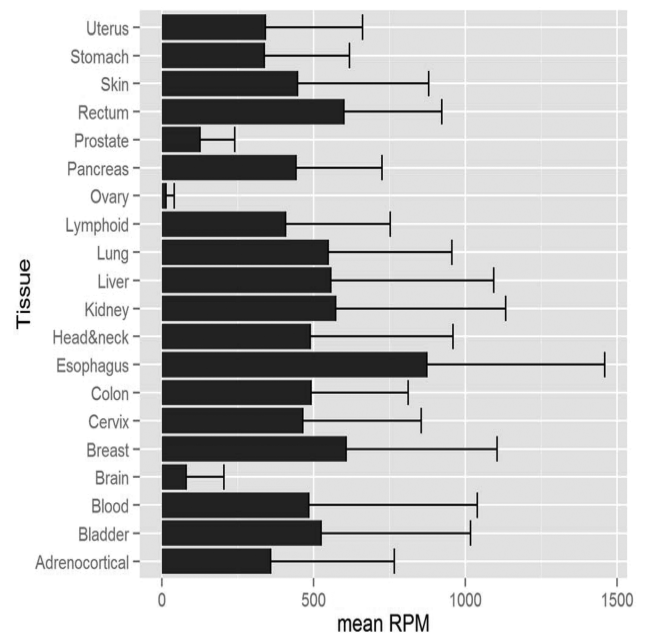


Figure 3. Two new charts for isomiRs. Panel (A) shows the percentage of samples that expressed the isomiR in various tissues, and panel (B) indicates the mean expression of a specific isomiR.

(Supplemental Figure S2). This result indicates that most sequences only appear in one sample. We assume that if a sequence does exist in humans, multiple reads from multiple independent experiments will appear to support its existence. Thus, to reduce the false-positive rate, only sequences that met the criterion of being supported by >100 reads in >10 independent datasets were used for novel miRNA prediction according to the pipeline described in our previous study (15). In brief, the pooled dataset was used for novel miRNA prediction by miRDeep2 (19), mireap (20) and miRanalyzer (21). After unifying the prediction results of the three algorithms and filtering out novel miRNAs similar to known transcripts, 3467 putative novel miRNAs were predicted by at least one algorithm and 1408 of those were supported by CLIP-seq data (Supplemental Figure S3). The alignment results of both smRNA-seq and CLIP-seq are provided in the web interface for each putative novel miRNA (Figure 1).

META-ANALYSIS

YM500v2 adds a new function, Meta-analysis, which allows researchers to identify differentially expressed miRs and arm-switching events from two user-defined, specific sets of samples. We utilize an R/Bioconductor package, DESeq (22,23) and the algorithm described in our previous study (15) to identify differentially expressed miRs and arm-switching events, respectively. As shown in Supplemental Figure S4, users can select one or multiple datasets and define what sample type they would like to investigate, such as 'Primary Solid Tumor' or 'Solid Tissue Normal'. In addition, we also provide a list of clinical criteria, such as ICD-

O-3 histology, tumor stage, distant metastasis and lymph node status, to help researchers to select a subgroup of well-defined cancer samples according to one or multiple clinical parameters. After selecting two groups of samples, users could overview the detailed clinical information of the two groups of selected samples before submitting this job to the server for real-time calculation. The user would then receive a notification email with a Result ID, and could then see a visualization of the results (Figure 2) in the 'Result and Download' section when the job is completed.

ISOMIRS

Two statistical charts are added to the 'IsomiR' section to help researchers realize the expression profile of an isomiR across distinct tissues. miR-21-5p+CA, an isomiR of miR-21-5p adenylated by PAPD5 reported to lead the degradation of miR-21-5p (24), is used as an example (Figure 3). Figure 3A shows the percentage of samples that expressed this isomiR in various tissues, and Figure 3B indicates the mean expression of a specific isomiR.

DISCUSSION

The decreasing cost of sequencing technology has led to large amounts of small RNA sequencing data from cancer-related studies. To make the most of these valuable yet massive amounts of data, we have updated our YM500 database to make it focus on the human cancer miRNome. There are more than 8000 cancer-related samples incorporated into YM500v2, and a new function 'Meta-analysis', which

allows researchers to fully utilize these thousands of cancer samples, is also added. With the updated database, researchers will have the opportunity, for example, to identify a specific set of miRNAs for a distinct question if they could define samples according to a clarified biological or clinical goal. To achieve this, our 'Meta-analysis' function allows a user to select two groups of samples using dozens of clinical criteria and then identify differentially expressed miRNAs and arm-switching events for the two groups (Figure 2).

Putative novel miRNAs in YM500v2 were identified from thousands of samples, and more importantly, a new strategy was developed for filtering out false-positive discoveries. Compared with the previous version of YM500, YM500v2 has a lower number of novel miRNAs (from ~11 000 to ~3500), but has more novel miRNAs identified by three prediction algorithms (from 90 to 189). Moreover, 40.61% (1408/3467) of the putative novel miRNAs could be supported by CLIP-seq data, and a web interface displays the alignment results of both smRNA-seq and CLIP-seq.

Accumulating evidence suggests that a miRNA locus could produce a series of isomiRs during the miRNA biogenesis process (24–27). The phenomenon of isomiRs contributes largely to the dynamic miRNome and coinstantaneously presents a challenge for miRNA study. A few isomiRs have been proven to be functional regulatory molecules (24,27–31), and more studies about isomiRs should be performed and taken into consideration (32). Several *in silico* tools have been developed for identifying isomiRs from smRNA-seq data (33–36), and such tools will greatly promote the discovery of isomiRs and their relevant functional roles. YM500v2 provides the existing evidence of isomiRs from enormous smRNA-seq datasets for researchers whenever they discover significant isomiRs in their studies. A representative example of isomiRs is shown in Figure 3.

NGS has become the norm for large-scale cancer research, and cancer-related smRNA-seq data will accumulate rapidly in the future. YM500 will be updated periodically to incorporate new smRNA-seq data. We have developed a pipeline to process new data and to incorporate them into the database semiautomatically. Newly incorporated smRNA-seq data will also be re-annotated for meta-analysis. YM500 will continue to be an informative and valuable database for miRNA studies.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGEMENT

We are grateful to the National Center for High-performance Computing for computer time and facilities and thank the TCGA research network for the availability of data.

FUNDING

Ministry of Science and Technology (MOST) [101-2320-B-010-059-MY3, 102-2314-B-010-045, 102-2321-B-010-010, 103-2622-B-010-002]; National Health Research Institutes

[NHRI-EX102-10254SI]; Veterans General Hospitals University System of Taiwan (VGHUST) Joint Research Program; Tsou's Foundation [VGHUST102-G7-3-2]; UST-UCSD International Center for Excellence in Advanced Bioengineering sponsored by the Taiwan NSC I-RiCE Program [102-2911-I-009-101]; National Yang-Ming University (Ministry of Education, Aim for the Top University Plan). Funding for open access charge: MOST [101-2320-B-010-059-MY3, 102-2314-B-010-045, 102-2321-B-010-010, 103-2622-B-010-002]; National Health Research Institutes [NHRI-EX102-10254SI]; VGHUST Joint Research Program; Tsou's Foundation [VGHUST102-G7-3-2]; UST-UCSD International Center for Excellence in Advanced Bioengineering sponsored by the Taiwan NSC I-RiCE Program [102-2911-I-009-101]; National Yang-Ming University (Ministry of Education, Aim for the Top University Plan).

Conflict of interest statement. None declared.

REFERENCES

- Mendell, J.T. and Olson, E.N. (2012) MicroRNAs in stress signaling and human disease. *Cell*, **148**, 1172–1187.
- Friedman, R.C., Farh, K.K., Burge, C.B. and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
- Adams, B.D., Kasinski, A.L. and Slack, F.J. (2014) Aberrant regulation and function of microRNAs in cancer. *Curr. Biol.*, **24**, R762–R776.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
- Liu, C., Kelnar, K., Vlassov, A.V., Brown, D., Wang, J. and Tang, D.G. (2012) Distinct microRNA expression profiles in prostate cancer stem/progenitor cells and tumor-suppressive functions of let-7. *Cancer Res.*, **72**, 3393–3404.
- Andorfer, C.A., Necela, B.M., Thompson, E.A. and Perez, E.A. (2011) MicroRNA signatures: clinical biomarkers for the diagnosis and treatment of breast cancer. *Trends Mol. Med.*, **17**, 313–319.
- Farazi, T.A., Ten Hoeve, J.J., Brown, M., Mihailovic, A., Horlings, H.M., van de Vijver, M.J., Tuschl, T. and Wessels, L.F. (2014) Identification of distinct miRNA target regulation between breast cancer molecular subtypes using AGO2-PAR-CLIP and patient datasets. *Genome Biol.*, **15**, R9.
- Garzon, R., Volinia, S., Liu, C.G., Fernandez-Cymering, C., Palumbo, T., Pichiorri, F., Fabbri, M., Coombes, K., Alder, H., Nakamura, T. *et al.* (2008) MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood*, **111**, 3183–3189.
- Hayes, J., Peruzzi, P.P. and Lawler, S. (2014) MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol. Med.*, **20**, 460–469.
- Yin, H., Kanasty, R.L., Eltoukhy, A.A., Vegas, A.J., Dorkin, J.R. and Anderson, D.G. (2014) Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.*, **15**, 541–555.
- van Rooij, E. and Kauppinen, S. (2014) Development of microRNA therapeutics is coming of age. *EMBO Mol. Med.*, **6**, 851–864.
- Li, Z. and Rana, T.M. (2014) Therapeutic targeting of microRNAs: current status and future challenges. *Nat. Rev. Drug Discov.*, **13**, 622–638.
- Bader, A.G., Brown, D. and Winkler, M. (2010) The promise of microRNA replacement therapy. *Cancer Res.*, **70**, 7027–7030.
- Bouchie, A. (2013) First microRNA mimic enters clinic. *Nat. Biotechnol.*, **31**, 577–577.
- Cheng, W.C., Chung, I.F., Huang, T.S., Chang, S.T., Sun, H.J., Tsai, C.F., Liang, M.L., Wong, T.T. and Wang, H.W. (2013) YM500: a small RNA sequencing (smRNA-seq) database for microRNA research. *Nucleic Acids Res.*, **41**, D285–D294.
- Friedlander, M.R., Lizano, E., Houben, A.J., Bezdan, D., Banez-Coronel, M., Kudla, G., Mateu-Huertas, E., Kagerbauer, B.,

- Gonzalez, J., Chen, K.C. *et al.* (2014) Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol.*, **15**, R57.
17. Ladewig, E., Okamura, K., Flynt, A.S., Westholm, J.O. and Lai, E.C. (2012) Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res.*, **22**, 1634–1645.
 18. Friedlander, M.R., Mackowiak, S.D., Li, N., Chen, W. and Rajewsky, N. (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.*, **40**, 37–52.
 19. Friedlander, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S. and Rajewsky, N. (2008) Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.*, **26**, 407–415.
 20. Chen, X., Li, Q., Wang, J., Guo, X., Jiang, X., Ren, Z., Weng, C., Sun, G., Wang, X., Liu, Y. *et al.* (2009) Identification and characterization of novel amphioxus microRNAs by Solexa sequencing. *Genome Biol.*, **10**, R78.
 21. Hackenberg, M., Sturm, M., Langenberger, D., Falcon-Perez, J.M. and Aransay, A.M. (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res.*, **37**, W68–W76.
 22. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
 23. Anders, S., McCarthy, D.J., Chen, Y., Okoniewski, M., Smyth, G.K., Huber, W. and Robinson, M.D. (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.*, **8**, 1765–1786.
 24. Boele, J., Persson, H., Shin, J.W., Ishizu, Y., Newie, I.S., Sokilde, R., Hawkins, S.M., Coarfa, C., Ikeda, K., Takayama, K. *et al.* (2014) PAPD5-mediated 3' adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease. *Proc. Natl Acad. Sci. U.S.A.*, **111**, 11467–11472.
 25. Fukunaga, R., Han, B.W., Hung, J.H., Xu, J., Weng, Z. and Zamore, P.D. (2012) Dicer partner proteins tune the length of mature miRNAs in flies and mammals. *Cell*, **151**, 533–546.
 26. Wu, H., Ye, C., Ramirez, D. and Manjunath, N. (2009) Alternative processing of primary microRNA transcripts by Drosha generates 5' end variation of mature microRNA. *PLoS One*, **4**, e7566.
 27. Ma, H., Wu, Y., Choi, J.G. and Wu, H. (2013) Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proc. Natl Acad. Sci. U.S.A.*, **110**, 20687–20692.
 28. Hinton, A., Hunter, S.E., Afrikanova, I., Jones, G.A., Lopez, A.D., Fogel, G.B., Hayek, A. and King, C.C. (2014) sRNA-seq analysis of human embryonic stem cells and definitive endoderm reveal differentially expressed microRNAs and novel isomiRs with distinct targets. *Stem Cells*, **32**, 2360–2372.
 29. Tan, G.C., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I.M., Robinson, S., Zhang, S., Ellis, P., Langford, C.F. *et al.* (2014) 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.*, **42**, 9424–9435.
 30. Xia, J. and Zhang, W. (2014) A meta-analysis revealed insights into the sources, conservation and impact of microRNA 5'-isoforms in four model species. *Nucleic Acids Res.*, **42**, 1427–1441.
 31. Guo, L., Zhao, Y., Yang, S., Zhang, H. and Chen, F. (2014) A genome-wide screen for non-template nucleotides and isomiR repertoires in miRNAs indicates dynamic and versatile microRNAome. *Mol. Biol. Rep.*, **41**, 6649–6658.
 32. Guo, L. and Chen, F. (2014) A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene*, **544**, 1–7.
 33. Humphreys, D.T. and Suter, C.M. (2013) miRspring: a compact standalone research tool for analyzing miRNA-seq data. *Nucleic Acids Res.*, **41**, e147.
 34. Giurato, G., De Filippo, M.R., Rinaldi, A., Hashim, A., Nassa, G., Ravo, M., Rizzo, F., Tarallo, R. and Weisz, A. (2013) iMir: an integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-Seq. *BMC Bioinformatics*, **14**, 362.
 35. de Oliveira, L.F., Christoff, A.P. and Margis, R. (2013) isomiRID: a framework to identify microRNA isoforms. *Bioinformatics*, **29**, 2521–2523.
 36. Luo, G.Z., Yang, W., Ma, Y.K. and Wang, X.J. (2014) ISRNA: an integrative online toolkit for short reads from high-throughput sequencing data. *Bioinformatics*, **30**, 434–436.